The population genomic signature of environmental association and gene flow in an ecologically divergent tree species Metrosideros polymorpha (Myrtaceae)

Izuno, Ayako ; Kitayama, Kanehiro ; Onoda, Yusuke ; Tsujii, Yuki ; Hatakeyama, Masaomi ; Nagano, Atsushi J ; Honjo, Mie N ; Shimizu-Inatsugi, Rie ; Kudoh, Hiroshi ; Shimizu, Kentaro K ; Isagi, Yuji

Abstract: Genomewide markers enable us to study genetic differentiation within a species and the factors underlying it at a much higher resolution than before, which advances our understanding of adaptation in organisms. We investigated genomic divergence in Metrosideros polymorpha, a woody species that occupies a wide range of ecological habitats across the Hawaiian Islands and shows remarkable phenotypic variation. Using 1659 single nucleotide polymorphism (SNP) markers annotated with the genome assembly, we examined the population genetic structure and demographic history of nine populations across five elevations and two ages of substrates on Mauna Loa, the island of Hawaii. The nine populations were differentiated into two genetic clusters distributed on the lower and higher elevations and were largely admixed on the middle elevation. Demographic modelling revealed that the two genetic clusters have been maintained in the face of gene flow, and the effective population size of the high-altitude cluster was much smaller. A FST -based outlier search among the 1659 SNPs revealed that 34 SNPs (2.05%) were likely to be under divergent selection and the allele frequencies of 21 of them were associated with environmental changes along elevations, such as temperature and precipitation. This study shows a genomic mosaic of M. polymorpha, in which contrasting divergence patterns were found. While most genomic polymorphisms were shared among populations, a small fraction of the genome was significantly differentiated between populations in diverse environments and could be responsible for the dramatic adaptation to a wide range of environments.

DOI: https://doi.org/10.1111/mec.14016

DR. AYAKO IZUNO (Orcid ID : 0000-0002-9339-2845)

Received Date : 22-Sep-2015
Revised Date : 13-Dec-2016
Accepted Date : 28-Dec-2016
Article type : Original Article

The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha* (Myrtaceae)

Ayako Izuno[1, 2], Kanehiro Kitayama[1], Yusuke Onoda[1], Yuki Tsujii[1], Masaomi Hatakeyama [2, 3], Atsushi J. Nagano[4, 5, 6], Mie N. Honjo[6], Rie Shimizu-Inatsugi[2], Hiroshi Kudoh[6], Kentaro K. Shimizu[2, 6, 7], Yuji Isagi[1]

[1] Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
[2] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[3] Functional Genomics Center Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[4] Faculty of Agriculture, Ryukoku University, 1-5 Yokatani, Seta Ohe-cho, Otsu, Shiga 520-2194, Japan
[5] PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
[6] Center for Ecological Research, Kyoto University, 2-509-3 Hirano, Otsu, Shiga 520-2113, Japan
[7] Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka, Totsuka-ward, Yokohama, Kanagawa 244-0813, Japan

Key words: genome scan; genomic mosaics; Hawaii; *Metrosideros*; RAD sequencing

Corresponding author: Ayako Izuno
Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland (izunoay@gmail.com)

Running title: Population genomics of *Metrosideros*

ABSTRACT

Genome-wide markers enable us to study genetic differentiation within a species and the factors underlying it at a much higher resolution than before, which advances our understanding of adaptation in organisms. We investigated genomic divergence in *Metrosideros polymorpha*, a woody species that occupies a wide range of ecological habitats across the Hawaiian Islands and shows remarkable phenotypic variation. Using 1,659 single nucleotide polymorphism (SNP) markers annotated with the genome assembly, we examined the population genetic structure and demographic history of nine populations across five elevations and two ages of substrates on Mauna Loa, the island of Hawaii. The nine populations were differentiated into two genetic clusters distributed on the lower and higher elevations and were largely admixed on the middle elevation. Demographic modeling revealed that the two genetic clusters have been maintained in the face of gene flow and the effective population size of the high-altitude cluster was much smaller. A F_{ST}-based outlier search among the 1,659 SNPs revealed that 34 SNPs (2.05%) were likely to be under

divergent selection and the allele frequencies of 21 of them were associated with environmental changes along elevations, such as temperature and precipitation. This study shows a genomic mosaic of *M. polymorpha*, in which contrasting divergence patterns were found. While most genomic polymorphisms were shared among populations, a small fraction of the genome was significantly differentiated between populations in diverse environments and could be responsible for the dramatic adaptation to a wide range of environments.

INTRODUCTION

Genetic differentiation and the factors underlying it can provide relevant insights into processes that occur in genomes during early stages of speciation (Feder *et al.* 2012; Via 2012). Local adaptation by natural selection promotes genetic differentiation at specific loci responsible for adaptations while gene flow among populations prevents genetic differentiation at neutral regions. Population demography can also affect genetic differentiation across genomes. All these factors bring about heterogeneous differentiation in pattern and degree across a genome (i.e., genomic mosaics; Nosil *et al.* 2008; Hohenlohe *et al.* 2010; Feder *et al.* 2012; Gompert *et al.* 2012). Genomic mosaics can cause population differentiation sometimes resulting in reproductive isolation or speciation (Lexer *et al.* 2003;

Sweigart *et al.* 2006; Barr & Fishman 2010; Nosil & Schlüter 2011). Investigations of genomic mosaics are expected to demonstrate the extent of genetic changes that cause local adaptation and selective forces to promote genetic differentiation (Arnegard *et al.* 2014; Soria-Carrasco *et al.* 2014), which can reveal the genetic mechanisms underlying adaptations and predict the adaptive capacity of populations in response to environmental changes.

Rapid progress in high-throughput sequencing technologies has provided genome-wide information for various organisms at a lower cost than previous methods. Genome-wide markers have detected candidate genetic regions subject to selection using genome scanning or comparative genomics (e.g., Hohenlohe *et al.* 2010; Therkildsen *et al.* 2013; Baute *et al.* 2015). Even in neutral genetic regions, an increased number of markers can capture considerable sequence polymorphisms to enable the revelation of spatial genetic structure and demographic histories that could not have been obtained using a handful of neutral markers (e.g., Gutenkunst *et al.* 2009; Emerson *et al.* 2010; Hohenlohe *et al.* 2012; Wagner *et al.* 2012; Catchen *et al.* 2013; Excoffier *et al.* 2013).

The Hawaiian archipelago, which is approximately 4,000 km from the nearest continent, possesses extreme environmental gradients. Elevations rise to more than 4,000 m above sea level, and the annual mean temperature and precipitation range from 5 to 25°C and

from 250 to11,000 mm, respectively (Price 2004). Because of the continuous activity of volcanoes, soil ages also vary (Kitayama & Mueller-Dombois 1995). As a consequence of the typical characteristics found in island ecosystems, such as biogeographic isolation, rare colonization events, and released niches in an ecosystem, the flora of the Hawaiian Islands is composed of a limited number of species and is comparable to simple biomes with low diversity (Carlquist 1980). A species occupying a wide range of ecological niches in the Hawaiian Islands provides a means to study intraspecies genetic diversification, in which the evidence and trajectories of adaptive evolution of a single lineage are likely to be preserved.

Metrosideros polymorpha Gaud. (Myrtaceae), a tree species endemic to the Hawaiian Islands, is common on each island and occupies an array of ecological niches. It grows from sea level to the alpine timberline and exists in early-successional vegetation on new lava flows as well as in climax forests on mature substrates (Fig. 1a; Stemmermann 1983; Joel *et al.* 1994; Kitayama & Mueller-Dombois 1995; Cordell *et al.* 1998). Phenotypes such as leaf size, the amount of trichomes on leaf surfaces, and tree height are extremely variable across populations (Fig. 1b; Stemmermann 1983; Vitousek 1992; Kitayama *et al.* 1997). The leaf traits represented by leaf area, leaf mass per area, and trichome mass per area are highly predictable as a function of temperature, precipitation, and lava age (Tsujii *et al.*

2016). The remarkable phenotypic variation in this species has been recognized as differences between varieties. A total of eight varieties, which can be classified according to their habitat and leaf morphology, are known in the Hawaiian Islands and five of them are found on the island of Hawaii (Dawson & Stemmermann 1990). Previous studies targeting the 'pure' individuals of the five varieties on the island of Hawaii have shown the allele frequencies on microsatellite loci to be significantly differentiated between four hybridizing varieties, as one variety was not genetically distinguishable from the others (Stacy *et al.* 2014). The four varieties include *M. polymorpha* var. *polymorpha*, which is characterized by small, thick leaves with a large amount of trichomes and found in the subalpine zones (>1,800 m above sea level); *M. polymorpha* var. *glaberrima*, which has large, thin leaves without trichomes and mainly grows on the late-successional substrates at middle to lower elevations; *M. polymorpha* var. *incana*, whose leaf traits are intermediate between the above two varieties and mainly grows on early-successional substrates at the lower elevations; and *M. polymorpha* var. *newellii*, which is endemic to some waterways on the east side of the island of Hawaii and has glabrous, narrow-shaped leaves (Dawson & Stemmermann 1990; Stacy *et al.* 2014).

Besides the typical habitats of these four varieties, however, trees are also observed with intermediate morphologies that are not distinguishable as a particular variety, resulting in the continuous distribution of leaf traits along broad environmental gradients as well as within populations (Tsujii *et al.* 2016). In spite of the prediction that the continuous change in phenotypes could be due to frequent hybridization between varieties (DeBoer & Stacy 2013; Stacy *et al.* 2014; Stacy *et al.* 2016), a definition of variety and its admixture has not been supported by clustering of genotypes. Previous studies that used 9–10 microsatellite markers showed no clear genetic differences between individuals within a population or variety. The admixture proportions of the genetic clusters did not differ between individuals, populations, or varieties; thus, rather homogeneous spatial genetic structures across populations or varieties were detected (Harbaugh *et al*. 2009; Stacy *et al.* 2014). Therefore, varieties defined by morphological and ecological characteristics may not perform as actual evolutionary units and comparison of an array of individuals based on a clustering algorithm with a higher number of genetic markers can offer key insights into genetic differentiation within this species.

M. polymorpha likely existed on the island of Kauai, the oldest island of the current Hawaiian Islands, approximately 3.9 to 1 million years ago and subsequently colonized newer islands (Wright *et al*. 2001; Percy *et al.* 2008). During the colonization, this species might have experienced population bottlenecks resulting from founder effects or in response to climate oscillations between glacial and interglacial periods (Gavenda 1992), or experienced population growth in response to niche appropriations. The origin of ecological divergence within this species is also still unclear. On the island of Hawaii, a greater genetic differentiation between three common varieties (var. *polymorpha*, var. *glaberrima*, and var. *incana*) was found on an old volcano compared with young volcanoes, indicating that the ecological and genetic diversification of this species could originate from events before its colonization on the island of Hawaii (Stacy *et al.* 2014). Meanwhile, the endemism of one variety, var. *newellii*, implies the rapid evolution of ecological differentiation (Stacy *et al.* 2014). Estimating the time scale of the ecological divergence and the demographic history of the species is expected to advance our understanding of the current population genetic structure as well as the evolutionary history of ecological divergence.

In *M. polymorpha,* a similar magnitude of variation in leaf traits was observed in a common garden (Cordell *et al.* 1998; Martin *et al.* 2007; Tsujii *et al.* 2016), suggesting the phenotypic variation observed in the field is based on genetic differentiation. Indeed, moderate to strong heritability was estimated in the leaf traits of two varieties, var. *incana* and var. *glaberrima* (Stacy *et al.* 2016). Trees with different leaf morphology show different functional properties such as efficiency of light use, stomatal traits, and leaf composition (Martin *et al.* 2007; Hoof *et al.* 2008). More importantly, reciprocal transplantation of seedlings of two varieties, var. *incana* and var. *glaberrima*, at the natural habitats showed that each variety performed better at their original habitat type (Morrison & Stacy 2014). These studies indicate that the variation in phenotype of this species could have adaptive consequences. However, the molecular mechanisms responsible for the variation in phenotypes and adaptive diversifications, in other words, the genomic insights into the selection, have not been shown.

We aim to elucidate genomic differentiation and the factors underlying it in *M. polymorpha* occupying a wide range of ecological niches. To enhance the genome-wide

studies of *M. polymorpha* (2*n* = 22), we recently reported a genome assembly (Izuno *et al.* 2016). By sequencing two paired-end and nine mate-pair libraries (up to 24 kb insertion size), we obtained 304-Mbp genome sequences. Half of the assembly was covered by 19 scaffolds with >5 Mbp (scaffold N50 length of >5 Mbp), which would correspond to the majority of the 22 chromosome arms (two arms for each of the 11 haploid chromosomes). This high-quality assembly is adequate for various genome-wide analyses and could make *M. polymorpha* a model organism to study adaptive radiation in a species. In this study, more than 1,000 polymorphic DNA markers annotated with the assembled genome sequences (Izuno *et al.* 2016) were used to (1) reveal the population genetic structure of an array of individuals growing across an environmental gradient, (2) estimate the demographic history of the genetic clusters found in the study sites, and (3) detect outlier loci, which show nonneutral differentiation patterns and may play key roles in adaptation to various environments.

MATERIALS AND METHODS

Population sampling

In June 2013, leaves were collected from 72 *M. polymorpha* trees growing in the common garden at the Volcano Agriculture Station, University of Hawaii (Kitayama *et al.* 1997;

Cordell *et al.* 1998). The 72 individuals were grown from the seeds of different mother trees collected from nine populations distributed at five elevations (150, 700, 1,200, 1,800, and 2,400 m above sea level) and two lava flows (150 and 3,000 years old) on the east flank of Mauna Loa, the island of Hawaii (Table 1; Fig. 2). The source populations were distributed over almost the full environmental range of habitats of the three varieties, var. *polymorpha*, var. *glaberrima*, and var. *incana* (DeBoer & Stacy 2013; Stacy *et al*. 2014). The 72 samples covered a wide range of phenotypic variations, especially in leaf area and trichome weight (Table 1). The sample number of each population is variable (Table 1) because the survival rate of trees was different between source populations at the time of the sample collection, that is, 25 years after the common garden was established.

Genotyping by sequencing

For the 72 individual samples, total genomic DNA was extracted using the modified CTAB method (Murray & Thompson 1980) and digested with the restriction enzymes *Bgl*II and *Eco*RI. The digested DNA fragments and two adapters (*Bgl*II adapter and *Eco*RI adapter) were ligated. Digestion and ligation were performed simultaneously at 37°C for 16 h. The reaction mixture consisted of 20 ng of genomic DNA, 5 units of *Bgl*II (New England BioLabs,

Ipswich, MA, USA), 5 units of *Eco*RI-HF (NEB), 1 × NEB buffer2 (NEB), 1 × bovine serum albumin (BSA) (New England BioLabs), 0.2 μM *Bgl*II adapter, 0.2 μM *Eco*RI adapter, 1 mM ATP (Takara, Shiga, Japan), and 300 units of T4 DNA ligase (Enzymatics, Beverly, MA, USA). The ligation product was purified using AMPure XP (Beckman Coulter, Tokyo, Japan) according to the manufacturer's instructions. One-tenth of the purified DNA was used in the PCR enrichment with the DNA polymerase KOD-Plus-Neo (TOYOBO, Osaka, Japan). Sequences of the adaptors and primers used in this study are shown in Table S1 (Supporting information). PCR product fragments of approximately 320 bp were selected using E-Gel SizeSelect 2% (Life technologies, Waltham, MA, USA). Each restriction-site–associated DNA sequence (RAD-seq) library was uniquely barcoded, then pooled for sequencing. The first 49 bp of the fragments were sequenced in two lanes on a HiSeq 2500 (Illumina, San Diego, CA, USA) using TruSeq v3 chemistry at BGI Tech Solutions (Hong Kong, China). A total of 237,218,370 reads and an average of 3,294,700 reads per sample were obtained (DRA accession: DRA004253). The reads were mapped to the genome sequences of *M. polymorpha* (Izuno *et al.* 2016; DDBJ accession BCNH01000001–BCNH01036376) using Bowtie2 (ver. 2.2.4; Langmead & Salzberg 2012) with default parameter settings. Single nucleotide polymorphism (SNP) sites were identified in the BAM files using the *ref_map.pl* pipeline

implemented in Stacks (ver. 1.27; Catchen *et al.* 2011). We then selected biallelic SNP loci with more than five read counts per genotype. For population genomic analysis, we allowed SNP loci to have missing data in less than 50% of the individuals in each population. SNP loci with a minor allele frequency of less than 5% in all the 72 samples were removed. When the minor allele frequency of SNPs is too low, it is difficult to distinguish polymorphisms from sequencing errors; they are therefore uninformative in the analyses of population genetic structure and searches of loci putatively influenced by divergent selection (Roesti *et al.* 2012; Larson *et al*. 2013). Only a single SNP was selected per mapped RAD-seq locus to remove tightly linked SNP loci (Ferchaud & Hansen 2015). Selection of SNP loci was performed using the 'populations' command implemented in Stacks (ver. 1.27; Catchen *et al.* 2011) and VCFtools (ver. 0.1.11; Danecek *et al.* 2011).

Population genetic structure and demographic history

To evaluate genome-wide genetic diversity within populations, expected heterozygosity (H_E) was calculated based on the genotypes at all SNP loci using GenoDive (ver. 2.0b27; Meirmans & Tienderen 2004). The effect of elevation and lava age on expected heterozygosity was examined using a generalized linear model (GLM) in which the

responsible variable was expected heterozygosity and the explanatory variables were elevation and lava age. We compared the Akaike's information criterion (AIC) scores of the four candidate models with different combinations of explanatory variables, and selected the most appropriate model with the lowest AIC. These analyses were performed with the *betareg* package (ver. 3.0.5; Cribari-Neto & Zeileis 2010) and *MuMIn* package (ver. 1.15.6; Barton 2013) in R (ver. 3.2.3; R Core Team 2014). Spatial genetic structure was investigated by two methods, Bayesian clustering (STRUCTURE ver. 2.3.4; Pritchard *et al.* 2000) and a principal component analysis (PCA). In the STRUCTURE analysis, an admixture model that assumed correlated allele frequencies between populations was used. Ten replicate simulations were run for each K (K = 1–9), with 50,000 burn-in steps followed by 100,000 Markov chain Monte Carlo steps. The optimal K was inferred based on the delta K method (Evanno *et al.* 2005) implemented in STRUCTURE HARVESTER (ver. 0.6.94; Earl & vonHoldt 2012). Admixture proportions from replicate simulations at the optimal K were averaged using CLUMPP (ver. 1.1.2; Jakobsson & Rosenberg 2007). PCA of the genotypes was conducted using GenoDive (ver. 2.0b27; Meirmans & van Tienderen 2004). The effect of elevation on the PCA ordination was evaluated using a Pearson's product-moment correlation test between elevation and PC1 or PC2 scores. The difference of PC1 or PC2 scores between lava ages was

evaluated by *t*-test. These statistical analyses were performed using R (ver. 3.1.2; R Core Team 2014).

The demographic history was examined based on site frequency spectrum (SFS). In this analysis, to obtain overall minor allele frequency observed in populations, we included SNPs with less than 5% of minor allele frequency and excluded SNPs with missing data. Due to the absence of an outgroup to provide the ancestral state of the focal populations, the distribution of minor allele frequency (i.e., the 'folded SFS') was used. To infer the demographic history of the two genetic clusters found in Bayesian clustering (hereafter, the two genetic clusters are distinguished as L and H clusters, respectively; see below), we assessed a total of 10 models, which are composed of five possible scenarios with and without gene flow between two genetic clusters (Fig. 3). These include simple population differentiation models (i.e., isolation-with-migration for M1 and isolation-without-migration for M6) and population differentiation models including the event of size change before (M2 and M7) or after (M3–M5 and M8–M10) population differentiation. A size-changing event after population differentiation was allowed in both genetic clusters in M3 and M8, only in the H cluster in M4 and M9, and only in the L cluster in M5 and M10 (Fig. 3a). To examine

the effect of gene flow and historical changes of population sizes (i.e., bottleneck or population growth) on the present spatial genetic structure, gene flow between the two genetic clusters was allowed in M1–M5 but not in M6–M10 (Fig. 3a). In each demographic model, effective population sizes at each time point (N_A, N_B, N_{BL}, N_{BH}, N_L, N_H; Table 2), historical time points of demographic events (T_{DIV}, T_{CHG}, T_{LCHG}, T_{HCHG}; Table 2) and a migration rate between the two genetic clusters (m) were estimated using the composite-likelihood approach implemented in fastsimcoal2 (ver. 2.5.2.21; Excoffier *et al.* 2013). Because the dataset was limited to polymorphic SNPs, we could not estimate a genome-wide mutation rate to obtain absolute values for parameters. Therefore, in every model, N_L was fixed (1.00) and the relative values of each parameter to N_L were estimated. In each model, 20 independent runs, with 100,000 coalescent simulations and 10–40 cycles of the likelihood maximization algorithm, were performed. The best fit model was determined based on delta AIC and AIC weight; the 95% confidence interval (CI) of each parameter was estimated using 100 nonparametric bootstrapped datasets.

Detection and characterization of outlier loci

To detect highly differentiated SNPs that could be subject to divergent selection in the nine

populations, a genome scan based on the Bayesian method was carried out using Bayescan (ver. 2.1; Foll & Gaggiotti 2008). Bayescan calculates the Bayes factor (BF), which is the ratio of posterior probabilities of the natural selection and neutral models at a given locus, and judges whether a locus was under natural selection (Foll & Gaggiotti 2008; Nielsen *et al.* 2009). The analysis was carried out under the default parameter settings as follows: 20 pilot runs of 5,000 iterations and an additional 50,000 burn-in iterations, followed by 5,000 iterations with a thinning interval of 10. The prior odds were set to 10, indicating that a neutral differentiation is 10 times more likely than selection at a locus. Posterior odds (PO) represent BF. Loci with $\log_{10}$ PO >1, which indicates that natural selection is 10 times more likely than neutral differentiation at a locus, were identified as outlier SNPs (Nielsen *et al.* 2009; Milano *et al.* 2014). Outlier loci searches using Bayescan were also conducted between the two genetic clusters found in the Bayesian clustering (L and H clusters; see below). For each outlier SNP, the correlations between population allele frequency and environmental variables were examined using Bayenv2 (Günther & Coop 2013), in which a covariance structure among populations was considered. A covariance matrix was calculated using SNPs with no missing data for the 72 samples. Environmental variables included annual mean temperature, annual precipitation, and lava age of each population. The data for annual mean

temperature and annual precipitation were downloaded from WorldClim (BIO01 for annual mean temperature and BIO12 for annual precipitation; Hijmans *et al.* 2005; Table 1). After 100,000 iterations, BF ≥10 indicated a significant association.

The population differentiation patterns were compared between the entire and the outlier SNPs using three methods: pairwise F_{ST} values among populations, an analysis of molecular variance (AMOVA), and a phylogenic network. Pairwise F_{ST} values were calculated using GenoDive (ver. 2.0b27; Meirmans & van Tienderen 2004). AMOVA divided the whole genetic variance into two hierarchical categories of "among individuals within populations" and "among populations" using Arlequin (ver. 3.5.2.1; Excoffier & Lischer 2010). Phylogenetic analyses were performed using the neighbor-net method (Bryant & Moulton 2004) implemented in SplitsTree (ver. 4.13.1; Huson & Bryant 2006) based on uncorrected *p*-distances between individuals.

To estimate the biological functions of the genes linked to outlier SNPs, protein sequences were extracted for the putative genes located within 10 kb of the outlier SNPs and then used for homology search analysis. For each protein sequence, putatively homologous protein and its gene ontology (GO) terms were searched in the *Arabidopsis thaliana* protein sequence database provided by BLAST2GO (ver. 4.0.2; Conesa *et al.* 2005). Both the

BLASTP searching and GO annotations were conducted with an *E*-value threshold of 1.0×10^{-5}.

RESULTS

SNP calling

From a total of 237,218,370 RAD-seq reads obtained from 72 samples, 225,183,124 reads (95%) were successfully mapped to the reference genome sequences (Table S2, Supporting information). After the filtering of 9,695 SNPs that were identified with Stacks, 1,659 SNPs were used for the following population genomic analysis (Data S1, Supporting information). The SNPs were located on the 154 scaffolds of the genome assembly, but because of the fairly long scaffold length, we treated these SNPs as independent. The number of SNPs recovered from a sample was 1,538 on average (range, 906–1,640), and the number of samples sharing a given SNP was 67 on average (range, 28–72) (Fig. S1, Supporting information).

Population genetic structure and demographic history

Genetic diversity within populations and population genetic structure were evaluated using genotype data of the entire 1,659 SNPs. Expected heterozygosity in a population ranged from

0.19 to 0.25 (Fig. 4). The model selection based on AIC revealed that the best GLM model to predict the expected heterozygosity included elevation but not lava age. Elevation had a negative effect (estimated coefficient: $-1.252E^{-4} \pm 3.061E^{-5}$ SD) on expected heterozygosity (Fig. 4). In Bayesian clustering of the nine populations, delta *K* peaked at *K* = 2 followed by *K* = 3 (Fig. 5a). In the scenario of *K* = 2, a clear genetic differentiation between high and low elevation was found with partial admixture at the middle elevation (1,200 m) (Fig. 5b). The two genetic clusters mainly distributed in higher and lower elevations are defined as the H and L cluster, respectively. The *F* value, which represents the degree of genetic drift exposed during genetic differentiations from a common ancestry of estimated genetic clusters, of the L and H clusters was 0.09 and 0.22, respectively, and F_{ST} between the two genetic clusters was 0.03. Expected heterozygosity within the L and H clusters was 0.26 and 0.23, respectively. The L cluster was further differentiated into two clusters in the scenario of *K* = 3 (Fig. 5b). In this scenario, F_{ST} between genetic clusters was 0.04–0.05 and expected heterozygosity within a cluster was 0.23–0.24. PCA of genotypes at the 1,659 SNPs also revealed the primary axis of differentiation to be among elevations; the PC1 axis, which explained 9.70% of the total genetic variance (Fig. 5c), showed a significant correlation with elevation ($r = 0.86$; $p < 0.001$). A PC2 axis with 5.77% of total genetic variance did not show significant correlation

with either elevation or lava age. However, PC2 scores of 56 samples in the six populations at 150, 700, and 1,200 m were significantly different between lava ages ($p = 0.05$) (Fig. 5c).

The demographic histories of the two genetic clusters found in the Bayesian clustering analysis (L and H clusters; Fig. 5b) were inferred by SFS. Because the L and H clusters were largely admixed in the populations at 1,200 m, a two-dimensional (2-D) SFS between the L and H clusters was calculated using 36 samples from four populations at 150 and 700 m (i.e., O150, Y150, O700, and Y700) and 16 samples from three populations at 1,800 and 2,400 m (i.e., O1800, Y1800, and O2400). By discarding the SNPs with missing data in the 52 samples and retaining any SNPs regardless of a threshold of minor allele frequency, 1,830 SNPs (Data S2, Supporting information) were used to calculate the 2-D SFS (Fig. S2, Supporting information). The model with the lowest AIC was M5 (Fig. 3b, c). In the M5 model, an ancestral population diverged at $1.01E^{-2}$ (95% CI, $6.98E^{-3}$–$2.26E^{-2}$) × N_L generations before the present and ancestral L cluster with an effective population size of 0.06 × N_L expanded approximately 16 times at $2.93E^{-4}$ (95% CI, $1.52E^{-4}$–$1.20E^{-3}$) × N_L generations before present, where N_L indicates the effective population size of the current L cluster (Table 2). In contrast to the remarkable population growth found in the L cluster, the H cluster has likely kept its small effective population size of 0.03 × N_L since the event of

population differentiation. Similar demographic parameters were also estimated in M3 with the second lowest AIC (Fig. 3b, c), in which the cluster H additionally increased effective population size at $6.70E^{-3} \times N_L$ generations before present (Table 2). The large AIC scores and near zero AIC weight in the M6–M10 models (Fig. 3b, c) indicate the importance of gene flow between the two genetic clusters in the demographic history of this species. In Table 2, the values of N_L are fixed (1.00) and the relative values of each parameter to N_L are shown; in the model of M5, for example, when N_L is assumed to be 100,000, each demographic parameter is indicated as follows: N_A = 1,000, N_{BL} = 600, N_H = 300 (individuals), T_{DIV} = 1,010, T_{LCHG} = 29.3 (generations before present), and $m = 5.03E^{-6}$ (individuals/generation).

Identification and characterization of outlier loci

Of the 1,659 SNPs, 34 (2.05%) on 21 scaffolds had $\log_{10}$ PO >1 in Bayescan, which suggests a divergent differentiation between the nine populations (Fig. 6a). Outlier loci searches conducted between the L and H clusters (36 and 16 samples, respectively) confirmed significant differentiation ($\log_{10}$ PO ≥1) at five outlier SNPs (Outliers 01, 10, 15, 29, and 30). The average F_{ST} value per SNP estimated in Bayescan was 0.15 (range, 0.05–0.51) (Fig. 6b). More than 95% of the SNPs (1,578 of 1,659; 95.8%) showed $F_{ST} < 0.2$ (Fig. 6b). The change

in allele frequencies at the 34 outlier SNPs among populations was 0.93 on average (range, 0.64–1.00) (Fig. 7), which suggests that the populations were indeed greatly differentiated at the outlier SNPs. We further tested the association between population allele frequency at the 34 outlier SNPs and three environmental variables. We found that 21 SNPs showed significant association (BF ≥10) with annual precipitation. Among the 21 SNPs, 16 were associated with annual mean temperature in addition to annual precipitation, possibly because of the correlation between these two environmental variables (Fig. 7). No outlier SNPs had an association with lava age. The F_{ST} value between the nine populations was 0.03–0.22 at the entire set of SNPs and 0.04–0.81 at the outlier SNPs (Table S3, Supporting information). AMOVA revealed the proportion of genetic variance among and within populations was 37.72% and 62.28%, respectively, at the outlier SNPs in contrast to 11.16% and 88.84% at the entire set of SNPs (Fig. 8a, b). These patterns were supported by the phylogenetic networks, as that of the outlier SNPs had longer branch lengths between populations than that of the entire set of SNPs (Fig. 8a and b). These results suggest that genetic differentiations at the outlier loci contribute to the genetic differentiations between the nine populations.

In total, 97 putative genes were found within 10 kb (range, 128–9,983 bases) of the outlier SNPs. In BLASTP and GO analysis, 57 genes obtained significant hits (*E*-value < 1.0

$\times 10^{-5}$) against the protein sequences of *A. thaliana* (Table S4, Supporting information).

DISCUSSION

Spatial genetic structure and demographic background revealed by genome-wide SNP markers

The present genome-wide SNP markers revealed the spatial genetic structure across an array of individuals of *M. polymorpha* on the east slope of Mauna Loa. The nine *M. polymorpha* populations were principally differentiated by elevation (Fig. 5b and 5c). The Bayesian clustering clearly showed two genetic clusters, occupying the lower and higher elevations (Fig. 5b), and the coordinates on the PC1 axis significantly correlated with elevation (Fig. 5c). However, the spatial genetic structure of the nine populations was not likely determined by elevation alone. In the Bayesian clustering, the scenario of $K = 3$ showed the second largest value of delta K compared with the other scenarios (Fig. 5a), indicating the presence of an additional genetic cluster besides the two detected in the scenario of $K = 2$. The PC1 axis explained only 9.05% of the genetic variance (Fig. 5c), suggesting that other factors

in addition to elevation also affected the spatial genetic differentiation. One of the candidate factors is lava age as PC2 scores of the trees in the lower elevations (150, 700, and 1,200 m) alone were significantly differentiated between lava ages.

The scenario of $K = 3$ in the Bayesian clustering (Fig. 5b) and the three-way structure in the PCA (Fig. 5c) likely represent the genetic differentiation between the three common varieties of *M. polymorpha* var. *polymorpha*, var. *glaberrima* and var. *incana*. The clear genetic differentiation between a genetic cluster occupying the higher elevations regardless of lava age (orange in Fig. 5b) and the other two clusters distributed in the lower elevations (blue and green in Fig. 5b) was found in both the scenario of $K = 2$ and $K = 3$, probably corresponding to a greater differentiation between var. *polymorpha* and the other two varieties (var. *glaberrima* and var. *incana)* (DeBoer & Stacy 2013; Stacy *et al*. 2014). The genetic cluster distributed mainly in young and old lava flow in the lower elevations with a large admixture likely correspond to var. *incana* and var. *glaberrima*, respectively. The relatively high F_{ST} between Y150 and O700 (Table S3, Supporting information), which represents typical habitats of var. *incana* and var. *glaberrima*, is consistent with the genetic and ecological differentiation between the two varieties (DeBoer & Stacy 2013; Morrison &

Stacy 2014; Stacy *et al.* 2016).

In contrast, as many as seven of the nine populations (45 of 72 trees) turned out to be a mixture of clusters (K = 3 in Fig. 5b), in which the assignment of a population or an individual to a particular variety was not feasible. The PCA plot indicates that the genetic differentiation across/within populations was fairly continuous (Fig. 5c) and only 11.16% of genetic variance explained the population differentiation (Fig. 8a). These large admixtures across/within populations can be attributed to frequent gene flow among populations, as the demographic modeling estimated the presence of gene flow between genetic clusters even after differentiation (Table 2; Fig. 3b and 3c). It should be noted that, in the demographic simulation, populations with the admixture of clusters were excluded, meaning that the analysis probably underestimated gene flow. Admixture between genetic clusters likely occurs at the overlap in preferred habitat (DeBoer & Stacy 2013) due to frequent hybridizations (Corn & Hiesey 1973; Stacy *et al*. 2016), although it sometimes occurs over a wide geographic distance as shown by an individual sample partially assigned to the L cluster on the highest elevation (Fig. 5b). *M. polymorpha* is bird and insect pollinated (Carpenter 1976; Koch & Sahli 2013) and produces a large number of seeds dispersed by the wind (Drake 1992). In addition to the high dispersal ability of this species, the absence of

geographic barriers (Aradhya *et al*. 1993) could lead to high gene flow among populations across a wide range of environmental conditions. Besides gene flow, shared ancestral polymorphism could also contribute to small genetic differentiations between/within populations.

Demographic modeling estimated contrasting effective population sizes of the two genetic clusters. The L cluster likely experienced a population expansion (M5 and M3 in Table 2). Because a larger population size is expected to increase population heterozygosity (Vila *et al*. 2003; Ortego *et al*. 2007), the population expansion event resulting in the larger population size presumably contributed to higher heterozygosity in the three populations on the lower to middle elevations (150, 700, and 1,200 m) (Fig. 4). On the other hand, the H cluster probably kept a relatively small population size after the event of population differentiation (M5 in Table 2). Even when considering M3, where a small population expansion in the H cluster was estimated, the historical population size of the H cluster is much smaller than that of the L cluster (Table 2). This bottleneck could explain a large differentiation of the H cluster from a common ancestry of the genetic clusters (F = 0.30; Fig. 5b) as well as the lower genetic variance within populations (Fig. 4). In assuming the gene flow between genetic clusters (M5 and M3 in Table 2), the low genetic diversity found in the

higher elevations could be also attributed to strong selection due to the harsh environmental conditions as shown in the low temperature and precipitation, limited nitrogen, and strong wind (Vitousek 1992; Stacy *et al*. 2014).

The population history shown in the present study still leaves questions about the evolutionary history of *M. polymorpha*. Since the present SNP dataset did not include monomorphic sites, which are required to convert the observed polymorphic rate to absolute time via a mutation rate (Excoffier *et al*. 2013), we could not determine absolute timings of the historical events. Further studies that include those with whole genome sequences of populations (e.g., Nadachowska-Brzyska *et al*. 2013; Meyer *et al*. 2016) or other types of genetic markers (Sousa & Hey 2013) are needed to estimate absolute timings of demographic events and to elucidate the origin of the polymorphic phenotypes and colonization history of this species.

Evidence of adaptive differentiation in *M. polymorpha* populations

A genome scan based on 1,659 SNP markers revealed that allele frequencies of the 34 SNPs significantly differed between populations (Fig. 6). Compared with the genetic

differentiations across the entire set of SNPs, the 72 samples strongly diverged at the outlier SNPs. Larger genetic distances between individuals were shown in the phylogenetic network of the outlier loci, although the differentiation pattern was almost identical (Fig. 8). Because observed allele frequencies at some of the outlier loci were significantly associated with environmental variables and linked to ecologically relevant genes (Table S4, Supporting information; discussed below), the genetic differentiations at the outlier loci are likely to be associated with environmental selection in this species. At the outlier loci, strong environmental selection could promote genetic divergence and overcome the homogenizing effects of gene flow among populations.

The 34 outlier SNPs could be considered a minimum set of candidate loci responsible for the environmental adaptations of this species. Focusing on a whole genome, instead of on a limited number of polymorphisms obtained with RAD-seq, can provide an increased number of candidate loci with more relevant information including linkage disequilibrium and strength of selections (e.g., Jones *et al*. 2012; The 1001 Genomes Consortium 2016). Moreover, although the nine populations covered almost the full ecological range of habitats of the three common varieties, they harbored a partial correlation

between geographic and environmental distances (Table 1). This correlation could affect the number of false positives in F_{ST} outlier tests, although the F_{ST} outlier test and association test used in this study took population structure into account. Selection scans across populations where geographic distance and environmental difference are less correlated would be more desirable (Wang & Bradburd 2014) and would provide comprehensive sets of candidate loci associated with environmental adaptations.

Within 10 kb of the 34 outlier SNPs, 97 putative genes were found (Table S4, Supporting information). Among the 34 SNPs, here we focused on the 21 SNPs at which allele frequency was associated with temperature or precipitation at the sites (Fig. S4 & S5, Supporting information). All the 21 SNPs were associated with precipitation (Fig. S5, Supporting information) and 16 SNPs were additionally associated with temperature (Fig. S4, Supporting information). Among the 21 precipitation-associated SNPs, Outliers 29, 23, 26, 15, and 09 are linked with g15450, g31360, g12854, g26807, and g22933, whose homologs in *A. thaliana* are involved in water deprivation, physiological functions stimulated by abscisic acid including stomatal movement, and cuticle development (Table S4, Supporting information). The phospholipase D delta gene of *A. thaliana* (AT4G35790), which was predicted to be homologous to g26807, is known to function in abscisic acid signaling and stomatal closure in

response to a water defect (Distéfano *et al*. 2012; Uraji *et al*. 2012) and plays an important role in drought tolerance (Bargmann *et al*. 2009). E3 ubiquitin-protein ligase KEG (AT5G13530), which is homologous to g12854, also works in abscisic acid pathways (Stone *et al*. 2006; Liu & Stone 2013). Furthermore, Outlier 29, a SNP loci showing significant association with temperature (Fig. S4, Supporting information) linked with g15445 and g15449, probably functions in response to heat including cold (Table S4, Supporting information); the latter, g15449, was predicted to be homologous to the class I heat shock protein gene of *A. thaliana* (AT1G07400). Heat shock proteins that were encoded by the class I heat shock protein gene comprise large, highly ordered chaperone complexes that could protect denatured proteins under heat stress (Walters *et al*. 1996). This evidence indicates that the genetic differentiations at the outlier loci could reflect those of the genes relevant to environmental adaptations. Nevertheless, more evidence is needed to reveal the effects of genetic differentiation at candidate genes on the variation in leaf functional traits (i.e., leaf mass per area, the amount of trichomes, and foliar chemical compositions; Vitousek *et al.* 1990, 1992; Joel *et al.* 1994; Tsujii *et al.* 2016), physiological properties, (i.e., capacity for water transport and gas exchange; Cornwell *et al.* 2007), and life-history (Morrison & Stacy 2014) between different habitats. Expression analyses under controlled environmental

conditions (Fraser *et al.* 2015) or transgenic experiments using model plant species (Kobayashi *et al.* 2013) would be useful in examining the biological meaning of genetic differentiations in an ecological context.

Concluding remarks

The ecological divergence of *M. polymorpha* in the Hawaiian Islands shows a number of similarities to that of Hawaiian silverswords, which is one of the best studied examples of plant adaptive radiation (Schluter 2000; Lawton-Rauh *et al.* 2007). The silversword alliance occupies diverse ecological habitats along moisture gradients with a marked morphological and physiological diversity; however, little has been reported on reproductive isolation other than habitat differences. The group diverged 5.2 to 1.2 million years ago, and 30 species are recognized while an allozyme study showed little genetic differentiation (Witter & Carr 1988). Similarly, *M. polymorpha* is distributed across a wide range of environments: altitudes of 0–2,500 m, annual rainfall from 400 to more that 4,000 mm, and various soil ages. While this species has been considered a single species, it shows extreme phenotypic variation such that its variation in some leaf traits (e.g., leaf mass per area) is comparable to the global variation of evergreen woody species (Tsujii *et al.* 2016). Instead, the size and color of flowers do not

differentiate so much and pollinators are shared in a wide range of habitats within and across islands (Carpenter 1976). *Metrosideros polymorpha* also diverged relatively recently, 3.9 to 1 million years ago (Wright *et al.* 2001; Percy *et al.* 2008), and previous allozyme and microsatellite studies reported little to moderate genetic differentiation between individuals (Aradhya *et al.* 1993; Harbaugh *et al.* 2009; DeBoer & Stacy 2013; Stacy *et al.* 2014). We confirmed the low differentiation in most parts of the genome (11.16% of total genetic variance was found among populations; Fig. 8a) using genome-wide markers annotated with the genome assembly. More importantly, our genome-wide studies also revealed clear differentiation at a small number of nonneutral loci (34 of 1,659 SNPs; Fig. 6a) that were closely associated with differences in the environments of the habitats (Fig. 7). Given the shared polymorphisms across a genome, a small fraction of the genome could be subject to strong environmental selection that is responsible for the remarkable phenotypic divergence (Chamberlain *et al.* 2009; Vestergaard *et al.* 2015; Lamichhaney *et al.* 2016).

The processes of the adaptive radiation in silverswords and *M. polymorpha* in the Hawaiian islands may contrast with those of those of most previous studies in which niche separation and reproductive isolation evolved simultaneously by "magic traits" of ecological speciation, such as by flower color affecting the pollinator interaction of *Mimulus* and

Aquilegia and allopolyploidization affecting hydrological niches of *Cardamine* (Waser & Campbell 2004; Wu *et al.* 2008; Hodges & Derieg 2009; Nosil 2012; Shimizu-Inatsugi *et al.* 2016). We speculate that the highly heterogeneous environments in the Hawaiian Islands as well as a lack of competing species that exclude these species from some parts of the environments maintain the divergence of these plants even when reproductive isolation may not be strong. It could be considered that the *M. polymorpha* genome represent a very early stage of the ecological speciation scenario, where divergent selections on traits eventually lead to reproductive isolation between populations (Schluter 2001), therefore it could be used as a model system to study the initial roles of natural selection that induce ecological speciation.

ACKNOWLEDGMENTS

We thank Michimasa Yamasaki for his help with the statistical analyses and Ichiro Tamaki for his advice on the coalescent simulations. This work was supported by JSPS KAKENHI grants (22255002, 24248028, and 16H06469), a Grant-in-Aid for JSPS fellows (13J02516), the Swiss National Science Foundation, and University Research Priority Program of Evolution in Action of the University of Zurich.

REFERENCES

Aradhya KM, Mueller-Dombois D, Ranker TA (1993) Genetic structure and differentiation in *Metrosideros polymorpha* (Myrtaceae) along altitudinal gradients in Maui, Hawaii. *Genetical Research*, **61**, 159–170.

Arnegard ME, McGee MD, Matthews B *et al.* (2014) Genetics of ecological divergence during speciation. *Nature*, **511**, 307–311.

Bargmann BOR, Laxalt AM, Riet BT *et al.* (2009) Multiple PLDs required for high salinity and water deficit tolerance in plants. *Plant and Cell Physiology*, **50**, 78–89.

Barr CM, Fishman L (2010) The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. *Genetics*, **184**, 455–465.

Bartoń K (2013) *MuMIn: multi-model inference*.

Baute GJ, Kane NC, Grassa CJ, Lai Z, Rieseberg LH (2015) Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytologist*, **206**, 830–838.

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Carlquist S (1980) *Hawaii: A Natural History.* Pacific Tropical Botanical Garden, Lawai, Hawaii.

Carpenter FL (1976) Plant-pollinator interactions in Hawaii - pollination energetics of *Metrosideros collina* (Myrtaceae). *Ecology*, **57**, 1125–1144.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3*, **1**, 171–182.

Catchen J, Bassham S, Wilson T *et al.* (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, **22**, 2864–2883.

Chamberlain NL, Hill RI, Kapan DD, Gilbert LE, Kronforst MR (2009) Polymorphic Butterfly Reveals the Missing Link in Ecological Speciation. *Science*, **326**, 847–850.

Conesa A, Götz S, García-Gómez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Cordell S, Goldstein G, Mueller-Dombois D, Webb D, Vitousek PM (1998) Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. *Oecologia*, **113**, 188–196.

Cornwell WK, Bhaskar R, Sack L, Cordell S, Lunch CK (2007) Adjustment of structure and function of Hawaiian *Metrosideros polymorpha* at high vs. low precipitation. *Functional Ecology*, **21**, 1063–1071.

Corn CA, Hiesey WM (1973) Altitudinal variation in Hawaiian *Metrosideros*. *American Journal of Botany*, **60**, 991–1002.

Cribari-Neto F, Zeileis A (2010) Beta regression in R. *Journal of Statistical Software*, **34**, 1–24.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dawson JW, Stemmermann L (1990) Metrosideros (Gaud). In: *Manual of the flowering plants of Hawai'i* (eds Wagner WL, Herbst DR, Sohmer SH), pp. 964–970. Univ. Hawai'i Press, Honolulu.

DeBoer N, Stacy EA (2013) Divergence within and among 3 varieties of the endemic tree, 'Ohi'a Lehua (*Metrosideros polymorpha*) on the eastern slope of Hawai'i Island. *Journal of Heredity*, **104**, 449–458.

Distéfano AM, Scuffi D, García-Mata C, Lamattina L, Laxalt AM (2012) Phospholipase Dδ is involved in nitric oxide-induced stomatal closure. *Planta*, **236**, 1899–1907.

Drake DR (1992) Seed dispersal of *Metrosideros polymorpha* (Myrtaceae): a pioneer tree of Hawaiian lava flows. *American Journal of Botany,* **79**, 1224–1228.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16196–16200.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Dupanloup I, Huerta-SÃ nchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in*

Genetics, **28**, 342–350.

Ferchaud A-L, Hansen MM (2015) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: threespine sticklebacks in divergent environments. *Molecular Ecology*, **25**, 238–259.

Foll M, Gaggiotti O (2008) A Genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Fraser BA, Künstner A, Reznick DN, Dreyer C, Weigel D (2015) Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). *Molecular Ecology*, **24**, 389–408.

Gavenda RT (1992) Hawaiian Quaternary paleoenvironments: a review of geological, pedological, and botanical evidence. *Pacific Science*, **46**, 295–307.

Gompert Z, Lucas LK, Nice CC *et al.* (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Günther T, Lampei C, Barilar I, Schmid KJ (2016) Genomic and phenotypic differentiation of *Arabidopsis thaliana* long altitudinal gradients in the North Italian Alps. *Molecular Ecology*, 1–19.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

Harbaugh DT, Wagner WL, Percy DM, James HF, Fleischer RC (2009) Genetic structure of the polymorphic *Metrosideros* (Myrtaceae) complex in the Hawaiian Islands using nuclear microsatellite data. *Plos One*, **4**, e4698.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hodges SA, Derieg NJ (2009) Adaptive radiations: from field to genomic studies. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9947–9954.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical*

transactions of the Royal Society of London. Series B, Biological sciences, **367**, 395–408.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics*, **6**, e1000862.

Hoof J, Sack L, Webb DT, Nilsen ET (2008) Contrasting structure and function of pubescent and glabrous varieties of Hawaiian *Metrosideros polymorpha* (Myrtaceae) at high elevation. *Biotropica*, **40**, 113–118.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Izuno A, Hatakeyama M, Nishiyama T *et al.* (2016) Genome sequencing of *Metrosideros polymorpha* (Myrtaceae), a dominant species in various habitats in the Hawaiian Islands with remarkable phenotypic variations. *Journal of Plant Research*, **129**, 727–736.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Joel G, Aplet G, Vitousek PM (1994) Leaf morphology along environmental gradients in Hawaiian *Metrosideros polymorpha*. *Biotropica*, **26**, 17–22.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Kitayama K, Mueller-Dombois D (1995) Vegetation changes along gradients of long-term soil development in the Hawaiian montane rainforest zone. *Vegetatio*, **120**, 1–20.

Kitayama K, Pattison R, Cordell S, Webb D, Mueller-Dombois D (1997) Ecological and genetic implications of foliar polymorphism in *Metrosideros polymorpha* Gaud. (Myrtaceae) in a habitat matrix on Mauna Loa, Hawaii. *Annals of Botany*, **80**, 491–497.

Kobayashi MJ, Takeuchi Y, Kenta T *et al.* (2013) Mass flowering of the tropical tree *Shorea beccariana* was preceded by expression changes in flowering and drought-responsive genes. *Molecular Ecology*, **22**, 4767–4782.

Koch JB, Sahli HF (2013) Patterns of flower visitation across elevation and successional gradients in Hawai‘i 1. *Pacific Science*, **67**, 253–266.

Lamichhaney S, Fan G, Widemo F *et al.* (2016) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, **48**, 84–88.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Larson WA, Seeb LW, Everett MV *et al.* (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, **7**, 355–369.

Lawton-Rauh A, Robichaux RH, Purugganan MD (2007) Diversity and divergence patterns in regulatory genes suggest differential gene flow in recently derived species of the Hawaiian silversword alliance adaptive radiation (Asteraceae). *Molecular Ecology*, **16**, 3995–4013.

Lexer C, Lai Z, Rieseberg LH (2003) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist*, **161**, 225–233.

Liu H, Stone SL (2013) Cytoplasmic degradation of the *Arabidopsis* transcription factor ABSCISIC ACID INSENSITIVE 5 is mediated by the RING-type E3 Ligase KEEP ON GOING. *Journal of Biological Chemistry*, **288**, 20267–20279.

Martin RE, Asner GP, Sack L (2007) Genetic variation in leaf pigment, optical and photosynthetic function among diverse phenotypes of *Metrosideros polymorpha* grown in a common garden. *Oecologia*, **151**, 387–400.

Meirmans PG, van Tienderen PH (2004) Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Meyer RS, Choi JY, Sanches M *et al.* (2016) Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nature Genetics*, **48**, 1083–1088.

Milano I, Babbucci M, Cariani A *et al.* (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, **23**, 118–135.

Morrison KR, Stacy EA (2014) Intraspecific divergence and evolution of a life-history trade-off along a successional gradient in Hawaii's *Metrosideros polymorpha*. *Journal of Evolutionary Biology*, **27**, 1192–1204.

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, **8**, 4321–4325.

Nadachowska-Brzyska K, Burri R, Olason PI *et al.* (2013) Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLoS Genetics*, **9**, e1003942.

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus*

morhua). *BMC Evolutionary Biology*, **9**, 276–11.
Nosil P (2012) *Ecological speciation*. Oxford University Press, Oxford.
Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution*, **26**, 160–167.
Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.
Ortego J, Aparicio JM, Calabuig G, Cordero PJ (2007) Increase of heterozygosity in a growing population of lesser kestrels. *Biology Letters*, **3**, 585–588.
Percy DM, Garver AM, Wagner WL *et al.* (2008) Progressive island colonization and ancient origin of Hawaiian *Metrosideros* (Myrtaceae). *Proceedings of the Royal Society B: Biological Sciences*, **275**, 1479–1490.
Price JP (2004) Floristic biogeography of the Hawaiian Islands: influences of area, environment and paleogeography. *Journal of Biogeography*, **31**, 487–500.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
R Core Team (2014) R: A language and environment for statistical computing.
Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.
Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.
Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.
Shimizu-Inatsugi R, Terada A, Hirose K *et al.* (2016) Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Molecular Ecology*, 1–15.
Soria-Carrasco V, Gompert Z, Comeault AA (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, **344**, 738–742.
Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.
Stacy EA, Johansen JB, Sakishima T, Price DK, Pillon Y (2014) Incipient radiation within the dominant Hawaiian tree *Metrosideros polymorpha*. *Heredity*, **113**, 334–342.
Stacy EA, Johansen JB, Sakishima T, Price DK (2016) Genetic analysis of an ephemeral intraspecific hybrid zone in the hypervariable tree, *Metrosideros polymorpha*, on Hawai'i Island. *Heredity*, 1–11.

Stemmermann L (1983) Ecological studies of Hawaiian *Metrosideros* in a successional context. *Pacific Science*, **37**, 361–373.

Stone SL, Williams LA, Farmer LM, Vierstra RD, Callis J (2006) KEEP ON GOING, a RING E3 ligase essential for *Arabidopsis* growth and development, is involved in abscisic acid signaling. *The Plant Cell Online*, **18**, 3415–3428.

Sweigart AL, Fishman L, Willis JH (2006) A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics*, **172**, 2465–2479.

The 1001 Genomes Consortium (2016) 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 1–34.

Therkildsen NO, Hemmer-Hansen J, Als TD *et al.* (2013) Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. *Molecular Ecology*, **22**, 2424–2440.

Tsujii Y, Onoda Y, Izuno A, Isagi Y, Kitayama K (2016) A quantitative analysis of phenotypic variations of *Metrosideros polymorpha* within and across populations along environmental gradients on Mauna Loa, Hawaii. *Oecologia*, **180**, 1049–1059.

Uraji M, Katagiri T, Okuma E *et al.* (2012) Cooperative function of PLDδ and PLDα1 in abscisic acid-induced stomatal closure in *Arabidopsis*. *PLANT PHYSIOLOGY*, **159**, 450–460.

Vestergaard JS, Twomey E, Larsen R, Summers K, Nielsen R (2015) Number of genes controlling a quantitative trait in a hybrid zone of the aposematic frog *Ranitomeya imitator*. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20141950.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 451–460.

Vila C, Sundqvist AK, Flagstad O *et al.* (2003) Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 91–97.

Vitousek PM, Field CB, Matson PA (1990) Variation in foliar $\delta^{13}C$ in Hawaiian *Metrosideros polymorpha* - a case of internal resistance. *Oecologia*, **84**, 362–370.

Vitousek PM, Aplet G, Turner D, Lockwood JJ (1992) The Mauna-Loa environmental matrix - foliar and soil nutrients. *Oecologia*, **89**, 372–382.

Wagner CE, Keller I, Wittwer S *et al.* (2012) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria

cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, **23**, 5649–5662.
Waser NM, Campbell DR (2004) Ecological speciation in flowering plants. In: *Adaptive speciation:* (eds Dieckmann U, Doebeli M, Metz JAJ, Tautz D), pp. 264–277. Cambridge University Press, Cambridge.
Waters ER, Lee GJ, Vierling E (1996) Evolution, structure and function of the small heat shock proteins in plants. *Journal of experimental botany*, **47**, 325–338.
Witter MS, Carr GD (1988) Adaptive radiation and genetic differentiation in the Hawaiian Silversword alliance (Compositae, Madiinae). *Evolution*, **42**, 1278–1287.
Wright SD, Yong CG, Wichman SR, Dawson JW, Gardner RC (2001) Stepping stones to Hawaii: a trans- equatorial dispersal pathway for *Metrosideros* (Myrtaceae) inferred from nrDNA (ITS+ETS). *Journal of Biogeography*, **28**, 769–774.
Wu CA, Lowry DB, Cooley AM *et al.* (2008) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity*, **100**, 220–230.

DATA ACCESSIBILITY

- DNA sequences: DDBJ DRA accessions DRA004253
- SNP genotypes: uploaded as online Supporting information (Data S1 and Data S2)

AUTHOR CONTRIBUTIONS

A.I. and Y.I. designed the study; A.I., K.K., Y.O., Y.T. and Y.I. collected the samples; A.J.N., M.N.H. and H.K. conducted the library preparation for RAD sequencing; M.H., R.I.S. and K.K.S. provided the chance and took the procedures for genome sequencing in Functional Genomics Center Zurich; A.I. conducted bioinformatics, analyzed the data; A.I., K.K.S. and Y.I. wrote the manuscript with input from all coauthors.

Table 1 Summary of the 9 source populations of the 72 *M. polymorpha* individuals analyzed.

Population	Elevation (m)	Lava age (years)	Annual mean temperature (°C) *	Annual precipitation (mm) *	*N*	Leaf area (cm^2) (mean ± SD) †	Weight of trichome ($mg \cdot cm^{-2}$) (mean ± SD) †
O150	150	3000	22.6	3921	10	13.0 ± 5.5	1.2 ± 1.9
Y150	150	150	22.6	3862	7	10.9 ± 2.2	1.6 ± 0.9
O700	700	3000	18.7	3898	7	16.2 ± 5.1	1.0 ± 1.5
Y700	700	150	18.7	3898	12	12.6 ± 3.4	0.6 ± 2.4
O1200	1200	3000	15.7	2219	8	7.3 ± 1.2	1.7 ± 1.8
Y1200	1200	150	15.7	2219	12	7.2 ± 3.3	1.5 ± 2.2
O1800	1800	3000	13.0	1811	4	3.9 ± 1.1	8.3 ± 2.5
Y1800	1800	150	13.2	1983	6	4.9 ± 0.7	5.9 ± 1.5
O2400	2400	3000	10.3	1020	6	4.4 ± 1.4	7.6 ± 2.0

N, number of plant individuals analyzed; * Data from WorldClim (BIO01 and BIO12; Hijmans *et al.* 2005); †Data from Tsujii *et al.* (2016)

Table 2 Demographic modeling of the two genetic clusters found in the nine *M. polymorpha* populations on Mauna Loa. The best-fitted model was shown in bold.

Model	Maximum ln(Likelihood)	Number of parameters	AIC	Delta AIC	AIC weight	N_A	N_B	N_{BL}	N_{BH}	N_L	N_H	T_{DIV}	T_{CHG}	T_{LCHG}	T_{HCHG}	m
M1	-3044.096	5	6098.19	118.05	0.00	0.14				1.00	0.11	3.15E-02				1.19E-11
M2	-3047.380	7	6108.76	128.61	0.00	0.11	0.24			1.00	0.10	2.31E-02	3.18E-02			1.17E-11
M3	-2982.233	9	5982.47	2.32	0.24	0.02		0.05	0.01	1.00	0.03	9.71E-03		3.93E-04	6.70E-03	5.48E-11
M4	-3026.653	7	6067.31	87.16	0.00	0.15			0.02	1.00	0.18	3.86E-02			2.71E-02	1.14E-11
M5	**-2983.073**	**7**	**5980.15**	**0.00**	**0.76**	**0.01**		**0.06**		**1.00**	**0.03**	**1.01E-02**		**2.93E-04**		**5.03E-11**
M6	-3042.972	4	6093.94	113.80	0.00	0.23				1.00	0.24	2.89E-02				
M7	-3043.117	6	6098.23	118.09	0.00	0.21	0.30			1.00	0.26	3.02E-02	3.21E-02			
M8	-3038.319	8	6092.64	112.49	0.00	0.04		0.14	0.10	1.00	0.05	7.17E-03		1.82E-04	4.77E-03	
M9	-3041.250	6	6094.50	114.35	0.00	0.25			0.26	1.00	0.29	3.56E-02	2.86E-02			
M10	-3037.504	6	6087.01	106.86	0.00	0.07		0.21		1.00	0.08	9.13E-03		1.41E-04		

The two genetic clusters (L and H clusters) were distributed in the lower and higher elevations on the east flank of Mauna Loa (Fig. 5b). Effective population size in an ancestral population (N_A), an ancestral population after the event of population size change (N_B), the L and H cluster before the event of population size change (N_{BL} and N_{BH}), and current L and H cluster (N_L and N_H) were estimated (Fig. 3). Migration rate between the two genetic clusters (m), the time of population differentiation (T_{DIV}), and the time of population size changes in an ancestral population (T_{CHG}), the L (T_{LCHG}) and H (T_{HCHG}) were also estimated (unit: generations) (Fig. 3).

FIGURE CAPTIONS

Figure 1. (**a**) Various ecological habitats of *Metrosideros polymorpha*. (**b**) Remarkable variations in leaf morphology such as size, shape and color (reflecting presence or absence of trichome) of *M. polymorpha*. Leaves were collected in the common garden and abaxial side was shown for all leaves. An AA battery was included as a scale.

F**igure 2.** The location of the common garden at the Volcano Agriculture Station, University of Hawaii and the nine original seed sources of the 72 *Metrosideros polymorpha* trees analyzed in this study

Figure 3. (a) Schematic representation of demographic models. Demographic history of the two genetic clusters (L and H clusters) distributed in the lower and higher elevations on the east flank of Mauna Loa (Fig. 5b), was inferred based on the two-dimensional site allele frequency spectrum in fastsimcoal2 (Excoffier *et al.* 2013). The estimated parameters included effective population size in an ancestral population (N_A), an ancestral population after the event of population size change (N_B), the L and H cluster before the event of population size change (N_{BL} and N_{BH}), and current L and H cluster (N_L and N_H), and the time of population differentiation (T_{DIV}), and the time of population size changes in an ancestral population (T_{CHG}), the L (T_{LCHG}) and H (T_{HCHG}). Migration rate between the two genetic clusters (m) was included in M1–M5 but not in M6–M10. (b) Delta Akaike information criteria (AIC) score and (c) AIC weight for each model.

Figure 4. Population level heterozygosity in nine populations of *Metrosideros polymorpha* trees on Mauna Loa. The solid line indicates expected heterozygosity values predicted by a generalized linear model.

Figure 5. Spatial genetic admixture of 72 *Metrosideros polymorpha* trees across nine populations on Mauna Loa based on genotypes at the 1,659 single nucleotide polymorphisms (SNPs). (**a**) Plot of delta K as a function of the number of genetic clusters (K) according to Evanno *et al.* (2005). (**b**) Admixture proportions of genetic clusters in individual trees. Bar plots are shown for a scenario of $K = 2$ and $K = 3$. Population profiles for annual mean temperature and mean annual precipitation are also shown. (**c**) Principal component analysis

of genotypes at the 1,659 SNPs for the 72 *M. polymorpha* trees. Population profiles are shown in Table 1.

Figure 6. Results of Bayesian outlier analysis for 1,659 single nucleotide polymorphisms (SNPs) in *Metrosideros polymorpha*. (**a**) For each SNP, F_{ST} was plotted against $\log_{10}$ (posterior odds [POs]). SNPs with $\log_{10}$ (PO) >1 were recognized as outlier SNPs (shown in black). (**b**) Distribution of F_{ST} across the 1,659 SNPs.

Figure 7. Allele frequencies at the 34 outlier single nucleotide polymorphism loci in each of the nine populations of *Metrosideros polymorpha* on Mauna Loa. Black circles and gray squares indicate the allele frequencies in the populations on 3,000- and 150-year-old lava flows, respectively. A letter of T and P with outlier SNP ID indicates a significant association of allele frequency with annual mean temperature and annual precipitation, respectively, revealed by Bayenv2 (Günther & Coop 2013).

Figure 8. Phylogenetic network of the 72 individuals of *Metrosideros polymorpha* on Mauna Loa based on the *p*-distance at (**b**) genome-wide 1,659 and (**c**) the 34 outlier single nucleotide polymorphisms (SNPs). Inset pie charts indicate the proportion of genetic variance among and within populations calculated using analysis of molecular variance.

SUPPORTING INFORMATION

Data S1. Genotype data of 1,659 single nucleotide polymorphisms (SNPs) for 72 *Metrosideros polymorpha* samples (VCF format)

Data S2. Genotype data of 1,830 single nucleotide polymorphisms (SNPs) for 52 *Metrosideros polymorpha* samples used for the site-frequency-spectrum-based coalescent simulations. The 52 samples were from the 7 populations (O150, Y150, O700, Y700, O1800, Y1800, and 2400). (VCF format)

Table S1. Oligonucleotide sequences of adapters and primers used for RAD-seq libraries

Table S2. The number of raw RAD-seq reads and mapped reads in a draft genome sequence

for each individual

Table S3. Pairwise population differentiation (F_{ST}) among the nine *Metrosideros polymorpa* populations on Mauna Loa based on the genome-wide 1,659 (above the diagonal) and the 34 outlier (below the diagonal) SNP genotypes. Population profiles are shown in Table 1.

Table S4. Gene ontology results for the 97 putative genes located within 10 kb of the 34 outlier single nucleotide polymorphisms

Figure S1. Genotype data coverage of 1,659 single nucleotide polymorphisms (SNPs) among 72 *Metrosideros polymorpha* samples. (**a**) Distribution of the number of samples at an SNP site. Each SNP covered the genotype data for 67 samples, on average. (**b**) Distribution of the number of SNPs per sample. Each sample obtained the genotype data for 1,538 SNPs, on average.

Figure S2. Two-dimensional minor allele frequency spectrum between the two genetic clusters found in 72 *Metrosideros polymorpha* trees on Mauna Loa. The L and H clusters represent the genetic clusters distributed in the low (150 and 700 m) and high (1,800 and 2,400 m) elevation and defined based on the Bayesian clustering analysis (Fig. 5b).

Figure S3. Distribution of demographic parameters estimated by the 100 bootstrapped data for the model M5. Vertical lines indicate values estimated by the observed site frequency spectrum.

Figure S4. Correlation between allele frequency at the outlier loci and temperature of 9 populations. Only 16 outlier loci that have significant correlations with annual mean temperature in Bayenv2 were shown.

Figure S5. Correlation between allele frequency at the outlier loci and precipitation of 9 populations. Only 21 outlier loci that have significant correlations with annual precipitation in Bayenv2 were shown.